Supplementary Online Material (SOM)

The MspJI family of modification-dependent restriction endonucleases for epigenetic studies

Devora Cohen-Karni[1,2], Derrick Xu[2], Lynne Apone[2], Alexy Formenkov[2], Zhiyi Sun[2], Paul J. Davis[2], Shannon Kinney[2], Megumu Yamada-Mabuchi[2], Shuang-yong Xu[2], Theodore Davis[2], Sriharsa Pradhan[2], Richard J. Roberts[2], Yu Zheng[2,*]

1. MCBB program, Boston University, Boston, MA 02215
2. New England Biolabs Inc., 240 County Road, Ipswich, MA 01938

*****correspondence: zhengy@neb.com

Telephone: 1-978-380-7441

Cloning, protein expression and purification

The synthetic genes for MspJI, FspEI, LpnPI, AspBHI, RlaI and SgrTI were codon-optimized using in-house software for expression in *E. coli*. Cassettes of synthetic DNA (~500 bp) were first assembled using overlapping PCR from oligonucleotide DNA and then joined by USER cloning (1).

The synthetic MspJI, FspEI, LpnPI and SgrTI genes were each ligated into pTXB1 with an N-terminal His-tag and expressed in a *dcm*$^-$ *E. coli* strain T7 Express (C2566). Clones were grown in LB-Amp to OD_{600} 0.6~0.8 and induced with a final concentration of 0.5 mM IPTG. Induced cultures were then grown overnight at 25 ºC and stored as frozen cell pellets at -20 ºC. Each resuspended cell pellet was sonicated and the cleared lysate was collected after centrifugation. All purifications were carried out using an AKTA FPLC machine (GE Healthcare). MspJI was first purified on a HiTrap Heparin HP column, then a HisTrap HP column, and a final HiTrap SP column. FspEI, LpnPI and SgrTI were first purified on a HisTrap HP column, followed by a HiTrap Heparin HP column. The endonuclease activities of the fractions were assayed on regular λ DNA, which is partially *dcm* methylated. Each final protein product appeared as a single band on SDS-PAGE (Figure S1). The final concentrations of MspJI, FspEI and LpnPI after purification were estimated to be 1.3 mg/ml, 1 mg/ml and 2.3 mg/ml, respectively.

The synthetic RlaI and AspBHI genes were ligated into the vector pET21a and expressed in E. coli strain T7 Express (C2566). AspBHI endonuclease was purified by chromatography through Heparin DM, Bio-Gel HTP hydroxyapatite, Mono Q, and Heparin TSK columns. RlaI purification included an additional purification step on an HTP ceramic column following the Mono Q column. Each final protein product appeared as a single band on SDS-PAGE (Figure S1). The final concentrations of RlaI and AspBHI after purification were estimated to be 0.8 mg/ml, and 1 mg/ml respectively.

Unit definition of the enzymes

The cleavage efficiency of MspJI is enhanced by supplying short oligonucleotides with a methylated site, as previously shown (2). We found that this stimulation effect is a general property for the MspJI family of enzymes. Based on this, we define one unit as the amount of enzyme required to digest 1 µg of pBR322 (*dcm*$^+$) plasmid DNA to a stable pattern in the presence of 0.3 µM activator in one hour at 37ºC in a total volume of 50µl. The sequence and structure of the activator can be found in the Supplementary Materials (Figure S2).

All digestion reactions were carried out in standard NEB buffer 4 (50 mM potassium acetate, 20 mM Tris-acetate, 10 mM magnesium acetate, 1 mM dithiothreitol, pH 7.9 at 25ºC). pBR322 (*dcm*$^+$) was from NEB and pBR322 (*dcm*$^-$) was prepared using a *dam*$^-$/*dcm*$^-$ *E. coli* strain (#C2925, NEB).

T4αgt57βgt14 (3) and wild-type (hereafter T4 gt and T4 wt) genomic DNA were purified from phage cultures. 1 µg of DNA substrate was digested by 1 unit of each enzyme in the presence of 0.3 µM double-stranded DNA activator in a 50 µl volume. All reactions were incubated for 4 hours at 37 °C. The reaction products were resolved on a 1% agarose gel.

DNA oligonucleotides with or without internal methylated cytosines were synthesized either in-house or at Integrated DNA Technologies (IDT). In the digestion series used for sequence specificity determination, those used for cleavage site determination, and those to explore the consequences of multiple neighboring recognition elements (presented in Figure 2, Figure 3 and Figure S3), approximately 1 µM double-stranded oligonucleotides were digested by 1 unit of each enzyme in the absence or presence of 1.5 µM double-stranded DNA activator in a 10 µl volume. All reactions were incubated for 4 hours at 37°C. The reaction products were resolved using a 20% polyacrylamide 7M urea denaturing gel (Figure 2) or TBE native gel electrophoresis (Figure 3 and Figure S3).

Genomic DNA digestion

All genomic DNA digestion reactions were carried out in standard NEB buffer 4. Genomic DNA samples of HeLa and Jurkat cell DNA as well as 5-Aza-dC-treated (#N4003S) and enzymatically ^{m}CpG-methylated Jurkat cell DNA (#N4002S) were obtained from NEB, while other genomic DNAs were purchased from BioChain. [[Rabbit (Normal Rabbit Liver DNA Cat# D1834149, Lot #A802114), Corn (Normal plant Corn DNA Cat#D1634330, Lot #B307105) and Soy Bean (Normal plant Soy Bean DNA Cat# D1634370, Lot # B307107) DNA were purchased from BioChain (BioChain Institute, Inc. Hayward, CA)]]. 1 µg of DNA substrate was digested by 1 unit of each enzyme in the presence of 0.5 µM double-stranded DNA activator in a 30 µl volume at 37 °C for 16 hours. The reactions were then subjected to a 20% TBE native polyacrylamide gel electrophoresis (PAGE), and visualized by SYBR GOLD staining. Densitometry was performed on the Typhoon image using ImageQuant software (Figure 4B).

IMR90 Cell culture and DNA purification

IMR90 human lung fibroblasts were purchased from American Type Culture Collection (ATCC, Manassas, VA). Cells were maintained in Eagle's minimum essential medium with glutamine (EMEM; ATCC), 10% fetal bovine serum (Gemini BioProducts, West Sacramento, CA), and non-essential amino acids (ThermoFisher Scientific, Pittsburgh, PA). Cultures were maintained in a humidity controlled incubator at 37 °C with 5% CO2. Confluent cells were harvested by trypsinization for purification of genomic DNA. Genomic DNA was isolated using the Gentra Puregene Cell Kit (Qiagen, Valencia, CA)

as per the manufacturer's protocol. DNA purity and concentration were determined by Nanodrop before further analysis.

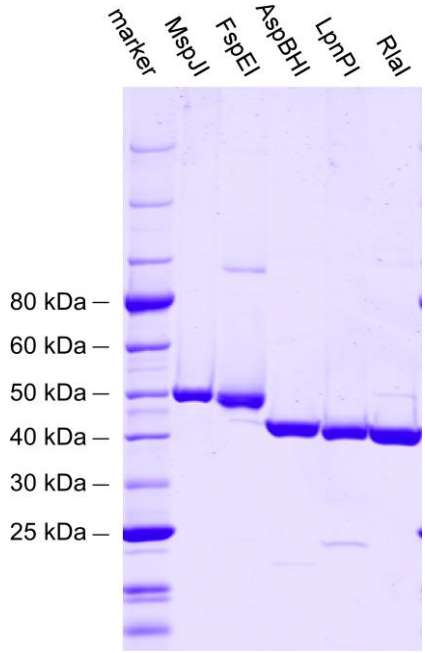
Sequencing - library construction

Purified IMR90 DNA was digested using MspJI (2 units of MspJI to each 1ug of genomic DNA) in the presence of activator (~0.5μM), and subjected to 10% TBE-PAGE gel. The 32-mer band was excised from the gel, and the DNA was extracted from the gel piece by soaking and centrifugation. Recovered DNA was used to construct sequencing library using the NEBNext ChIP-Seq kit which included DNA end-repair, adapters ligation and PCR amplification (NEBNext® ChIP-Seq Sample Prep Master Mix Set 1, E6240S). The final library was sequenced in-house using the SOLiD platform with 35 cycles according to the manufacturer's protocols.

Bioinformatic analysis

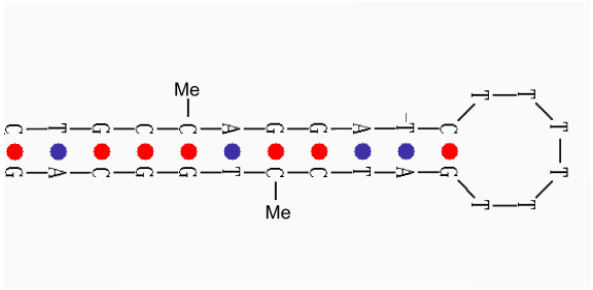
All 35-bp reads were aligned to the human reference (hg18) using bowtie (-n 0 –l 30 –e 500 –k 5 –best --strata) in sequence space (4). From the 5 best hits, the one which has the longest match to the reference is selected. If a read has multiple equally good hits, it is then discarded. Because of the general poor sequencing quality in the P2 adaptor region, identification of the P2 adaptor in the read is by heuristic rules. For example, if a read ends with AGA (the beginning sequence of the P2 adaptor) and has at least 1 mismatch in the last 3 bases, we consider it as a 32-mer. Or, if a read has the first mismatch at position 33, we consider it as a 32-mer. If all the 35 bases match to the reference, it is then considered to contain no less than 35-bp genomic fragment. Reads are classified into 30-35mer groups and sequence logos are constructed using weblogo (5).

For the subset of the Salk reference, we first chose fully methylated CG sites from the published IMR90 data (6). We then identified suitable MspJI recognition sites YNCGNR and extracted 32-mer sequences around these sites from the human reference genome. These 32-mers are then matched back to the human reference (hg18) using bowtie with option –m 1 to suppress multiple hits. The reported 32-mers are included into the reference subset.



SOM Figure S1. SDS-PAGE of the purified enzymes

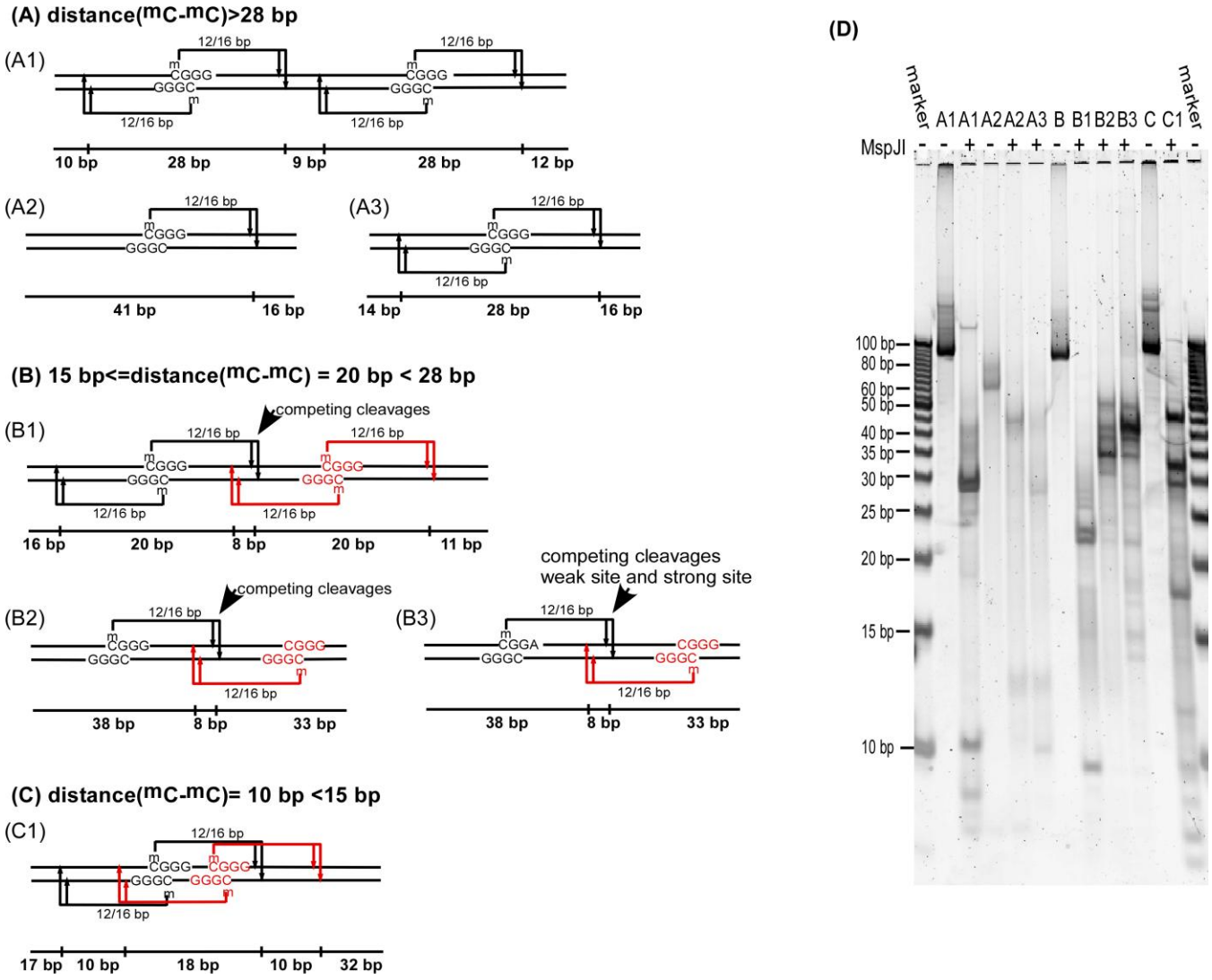
MspJI and homologs were over-expressed in *E. coli* and purified as described in Materials and Methods section. The purified proteins were resolved using a 4-20% gradient SDS-PAGE



SOM Figure S2. The activator

The activator is a short double-stranded oligonucleotide. It contains two methylated sites, but it is too short to allow cleavage.

The full activator sequence is: **CTGC^{m}CAGGATCTTTTTTGATC^{m}CTGGCAG** . The stem-loop structure it forms upon annealing is presented in this figure. The activator was previously found to stimulate MspJI digestion. The same effect was observed for all the homologous enzymes.



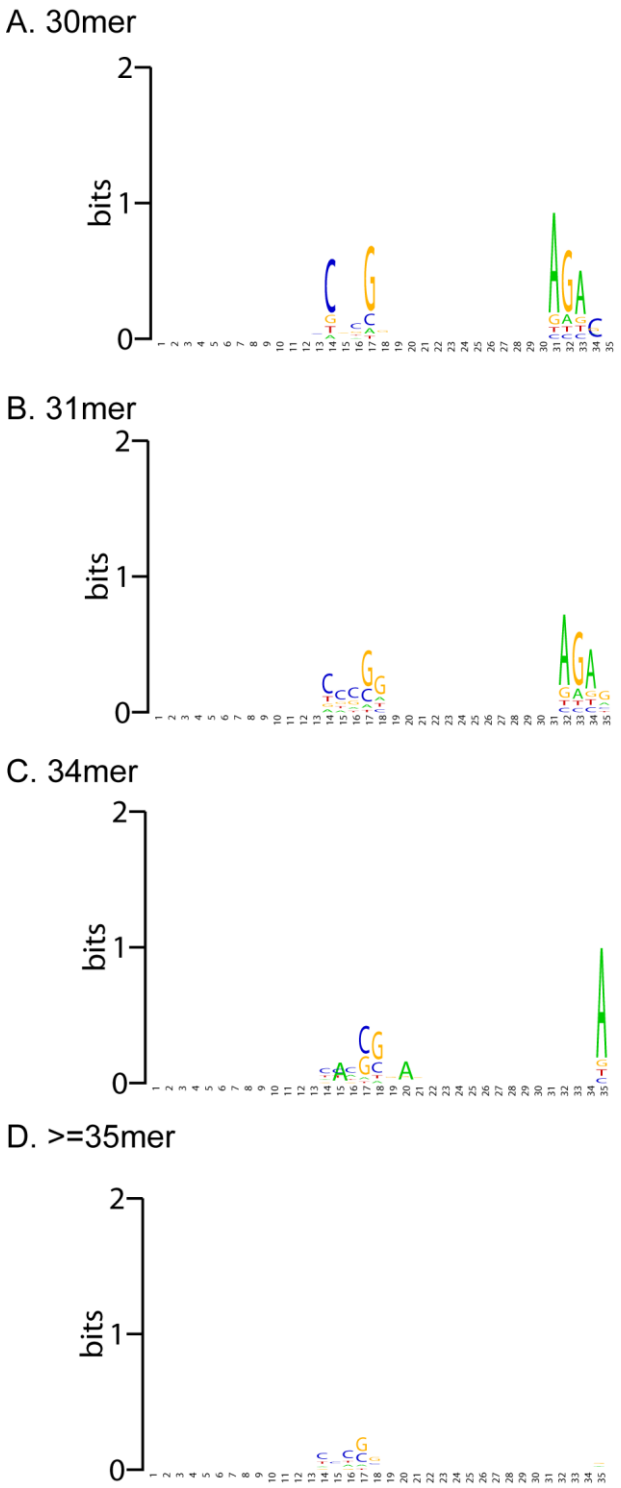
SOM Figure S3. Different MspJI cleavage scenarios

Experimental results for the possible cleavage scenarios are presented here. (**A**), (**B**) and (**C**) present the specific oligonucleotides labeled A1 – C1. The resulting digests, resolved on 20% TBE-PAGE are presented in (**D**).

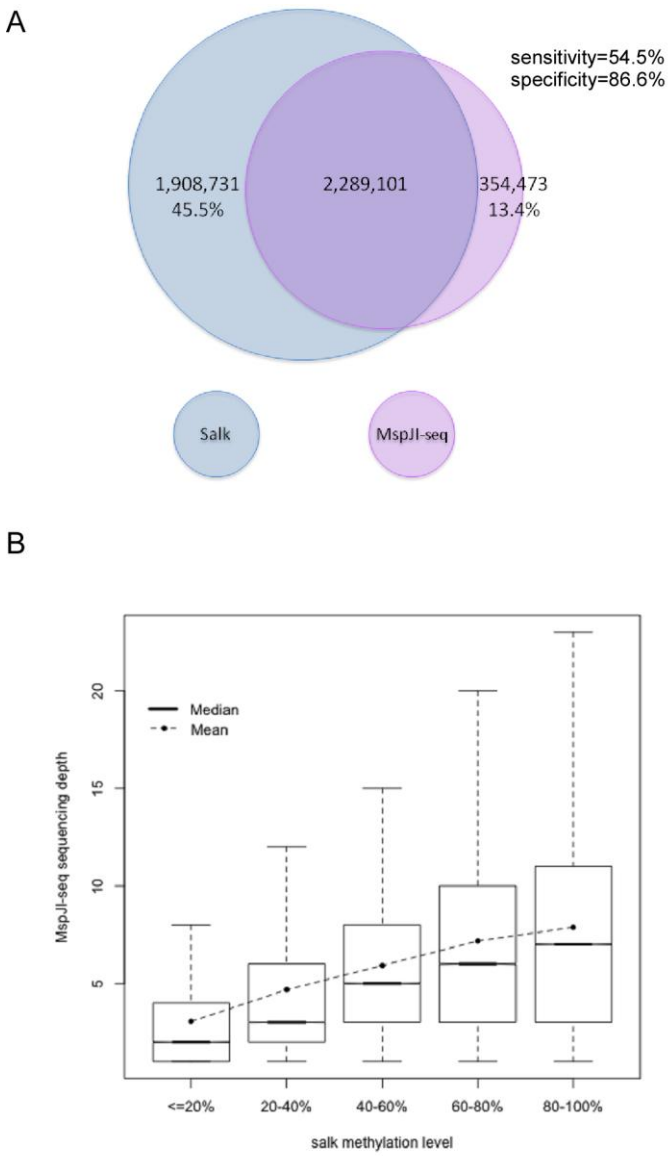
In genomic DNA, isolated recognition sites, where the distance between the closest modified cytosines (mC) is larger than 28 nucleotides, will be cleaved independently, either on both sides or on one side depending on the sequence context flanking the mC (Figure S3A). In the human genome, approximately 50% of all the CpG sites fall into this category (assuming they are all methylated). However, genomic DNA also will often contain modification at closely spaced sites, thus the cleavage pattern obtained in a

digestion using MspJI-like enzymes will depend on the distance between the neighboring recognition sites. Some examples of different disposition of nearby mC sites (designated “cleavage scenarios” for convenience) were explored in Figure S3. Recognition sites where the distance between the two mCs is less than 28 nucleotides but larger than 15 nucleotides, will affect each other’s cleavage pattern and will depend on the order of cleavage (Figure S3 lanes B1–B3). In such scenarios, fragments shorter than 32 nucleotides can be formed. For two recognition sites within 15 bases of each other, the final products also depend on the cleavage order (Figure S3 lane C1). More complex scenarios may arise when more than two neighboring recognition sites occur within 28 bases of each other. Thus, closely clustered ^{m}CpG sites as found in CpG islands may potentially reduce the amount of 32-mer produced during digestion.

SOM Figure S4 Aligned sequence homologs of MspJI.



SOM Figure S5. Sequence logo of the 30-mers (total 0.9M reads), 31mers (1.3M), 34-mers (0.7M), and >=35-mers (1.0M).



SOM Figure S6. **Comparison of the IMR90 MspJI-seq data and the Salk MethylC-seq data**

(A) Fully methylated CpG sites extracted from the pool of 32 and 33 bp reads of the MspJI-seq data were compared with a subset of the Salk MethylC-seq data which contain MspJI recognition sites. Three subgroups were identified: unique to the MspJI-seq, unique to the Salk subset and the overlap. The numbers in each subgroup and the percentages are also listed.

(B) Correlation between the MspJI-seq fold coverage and the Salk methylation level. Fully methylated $CN^{m}CGNG$ sites extracted from the pool of 32 and 33 bp reads of the MspJI-seq data were divided into 5 bins based on the Salk methylation level. The mean, median and overall distribution of the MspJI-seq fold coverage were illustrated by box plots and were compared across the 5 bins.

Set		Sequence
N^mcgN	**Top strand**	CGGCGTTTCCGGGTTCCATAGGCTCCGCC**N**^{m}CG**N**ACTCTGATGACCAGGGCATCACA
	Bottom strand	TGTGATGCCCTGGTCATCAGAGT**N**^{m}CG**N**GGCGGAGCCTATGGAACCCGGAAACGCCG
Nc^mcggN	**Top strand**	CGGCGTTTCCGGGTTCCATAGGCTCCGC**N**C^{m}CGG**N**CTCTGATGACCAGGGCATCACA
	Bottom strand	TGTGATGCCCTGGTCATCAGAG**N**C^{m}CGG**N**GCGGAGCCTATGGAACCCGGAAACGCCG
Ng^mcN	**Top strand**	CGGCGTTTCCGGGTTCCATAGGCTCCGCC**N**G^{m}C**N**ACTCTGATGACCAGGGCATCACA
	Bottom strand	TGTGATGCCCTGGTCATCAGAGT**N**G^{m}C**N**GGCGGAGCCTATGGAACCCGGAAACGCCG
N^mcNgN	**Top strand**	CGGCGTTTCCGGGTTCCATAGGCTCCGCC**N**^{m}C**N**G**N**ACTCTGATGACCAGGGCATCACA
	Bottom strand	TGTGATGCCCTGGTCATCAGAGT**N**^{m}C**N**G**N**GGCGGAGCCTATGGAACCCGGAAACGCCG
K^mcWHR	**Top strand**	CGGCGTTTCCGGGTTCCATAGGCTCCGCC**K**^{m}C**WHR**ACTCTGATGACCAGGGCATCACA

SOM Table S1. The full sequences of the oligonucleotides sets used for sequence specificity determination

N represents any nucleotide among A,T,C or G, not a randomized nucleotide mix

MspJI

set	N^mcgN			Nc^mcggN			Ng^mcN		N^mcNgN		N^mcNgN		K^mcWHR	
act.		-	+		-	+		+		+		+		+
	AcgA	+,-	+,-	AccggA	-	+,-	AgcA	-	AcAgA	+	TcAgC	-	TcACA	+
	TcgT	+	+	TccggT	-	-	TgcT	+	TcTgT	-	GcTgA	+	TcACG	+
	AcgT	+,-	+,-	AccggT	-	-	AgcT	-	AcTgA	+	TcTgC	-	TcAAA	+
	AcgT	+	+	AccggT	-	-	AgcT	+	TcAgT	-	GcAgA	+	TcAAG	+
	AcgC	+,-	+,-	AccggC	-	-	AgcC	-	AcCgA	+,-	TcCgC	-	TcATA	+
	GcgT	+	+	GccggT	-	-	GgcT	+	TcGgT	-	GcGgA	+	TcATG	+
	AcgG	+,-	+,-	AccggG	+	+	AgcG	-	AcGgA	+	TcGgC	-	GcACA	+,-
	CcgT	+	+	CccggT	-	-	CgcT	+	TcCgT	-	GcCgA	+	GcACG	+
	TcgA	-	+,-	TccggA	-	+,-	TgcA	-	AcAgT	-	TcAgG	N/T	GcAAA	+,-
	TcgA	+	+	TccggA	+	+	TgcA	+	AcTgT	-	CcTgA	N/T	GcAAG	+
	TcgC	+,-	+,-	TccggC	-	-	TgcC	-	AcCgT	-	TcTgG	+	GcATA	+
	GcgA	+	+	GccggA	-	+,-	GgcA	+	AcGgT	-	CcAgA	+	GcATG	+
	TcgG	+,-	+,-	TccggG	+	+	TgcG	-	AcAgC	-	TcCgG	+		
	CcgA	+	+	CccggA	+	+	CgcA	+	GcTgT	-	CcGgA	+		
	CcgC	+,-	+,-	CccggC	-	-	CgcC	-	AcTgC	-	TcGgG	+		
	GcgG	+	+	GccggG	+	+	GgcG	+	GcAgT	-	CcCgA	+		
	CcgG	+,-	+,-	CccggG	+	+	CgcG	-	AcCgC	-	CcAgC	-		
	CcgG	+	+	CccggG	+	+	CgcG	+	GcGgT	-	GcTgG	+		
	GcgC	+,-	+,-	GccggC	-	-	GgcC	-	AcGgC	-	CcTgC	-		
	GcgC	+	+	GccggC	-	-	GgcC	+	GcCgT	-	GcAgG	+		
									AcAgG	+	CcCgC	-		
									CcTgT	-	GcGgG	+		
									AcTgG	+	CcGgC	-		
									CcAgT	-	GcCgG	+		
									AcCgG	+	CcAgG	+		
									CcGgT	-	CcTgG	+		
									AcGgG	+	CcCgG	+		
									CcCgT	-	CcGgG	+		
									TcAgA	+	GcAgC	-		
									TcTgA	+	GcTgC	-		
									TcCgA	+	GcCgC	-		
									TcGgA	+	GcGgC	-		

SOM Table S2. Results of oligonucleotide digestion by MspJI

act. = activator; + = digested; - = undigested; +,- = partial digestion; N/T = not tested

FspEI

set	N^mcgN			Nc^mcggN			Ng^mcN			N^mcNgN			N^mcNgN			K^mcWHR		
act.		-	+		-	+		-	+		-	+		-	+		-	+
	AcgA	-	-	AccggA	+	+	AgcA	-	-	AcAgA	-	+,-	TcAgC	-	+	TcACA	-	+
	TcgT	-	-	TccggT	+	+	TgcT	-	-	TcTgT	-	+,-	GcTgA	-	+	TcACG	-	+
	AcgT	-	-	AccggT	+	+	AgcT	-	-	AcTgA	-	+,-	TcTgC	-	+	TcAAA	-	-
	AcgT	-	-	AccggT	+,-	+	AgcT	-	-	TcAgT	-	+,-	GcAgA	-	+	TcAAG	-	-
	AcgC	-	-	AccggC	+	+	AgcC	-	-	AcCgA	-	-	TcCgC	-	-	TcATA	-	-
	GcgT	-	-	GccggT	+	+	GgcT	-	-	TcGgT	-	+,-	GcGgA	-	+	TcATG	-	-
	AcgG	-	-	AccggG	+	+	AgcG	-	-	AcGgA	-	+,-	TcGgC	-	+	GcACA	-	+
	CcgT	+,-	+	CccggT	+,-	+	CgcT	-	-	TcCgT	-	-	GcCgA	-	-	GcACG	-	+
	TcgA	-	-	TccggA	+	+	TgcA	-	-	AcAgT	-	+	TcAgG	N/T	N/T	GcAAA	-	-
	TcgA	-	-	TccggA	+,-	+	TgcA	-	-	AcTgT	-	+,-	CcTgA	N/T	N/T	GcAAG	-	+
	TcgC	-	-	TccggC	+	+	TgcC	-	-	AcCgT	-	-	TcTgG	-	+,-	GcATA	-	-
	GcgA	-	-	GccggA	+,-	+	GgcA	-	-	AcGgT	-	+,-	CcAgA	+	+	GcATG	-	-
	TcgG	-	+,-	TccggG	+	+	TgcG	-	-	AcAgC	-	+	TcCgG	-	-			
	CcgA	+,-	+	CccggA	+,-	+	CgcA	-	-	GcTgT	-	+	CcGgA	+	+			
	CcgC	+	+	CccggC	+	+	CgcC	-	-	AcTgC	-	+,-	TcGgG	-	+			
	GcgG	-	-	GccggG	+	+	GgcG	-	-	GcAgT	-	+	CcCgA	+	+			
	CcgG	+	+	CccggG	+	+	CgcG	-	-	AcCgC	-	-	CcAgC	+	+			
	CcgG	+	+	CccggG	+,-	+	CgcG	-	-	GcGgT	-	+	GcTgG	-	-			
	GcgC	-	+,-	GccggC	+	+	GgcC	-	-	AcGgC	-	+	CcTgC	+	+			
	GcgC	-	-	GccggC	+,-	+	GgcC	-	-	GcCgT	-	-	GcAgG	-	-			
										AcAgG	+	+	CcCgC	+	+			
										CcTgT	+	+	GcGgG	-	-			
										AcTgG	-	-	CcGgC	N/T	N/T			
										CcAgT	+	+	GcCgG	N/T	N/T			
										AcCgG	-	-	CcAgG	+	+			
										CcGgT	+	+	CcTgG	+	+			
										AcGgG	-	+	CcCgG	+	+			
										CcCgT	+	+	CcGgG	+	+			
										TcAgA	-	+	GcAgC	-	+			
										TcTgA	-	+,-	GcTgC	-	+			
										TcCgA	-	-	GcCgC	-	-			
										TcGgA	-	+,-	GcGgC	-	+			

SOM Table S3. Results of oligonucleotide digestion by FspEI

act. = activator; + = digested; - = undigested; +,- = partial digestion; N/T = not tested

AspBHI

set	N^{m}cgN			Nc^{m}cggN			Ng^{m}cN			N^{m}cNgN			N^{m}cNgN			NccWggN	
act.		-	+		-	+		-	+		-	+		-	+		-
	AcgA	-	-	AccggA	-	-	AgcA	-	-	AcAgA	-	+,-	TcAgC	-	+,-	AccaggA	-
	TcgT	-	-	TccggT	-	+,-	TgcT	-	+,-	TcTgT	-	-	GcTgA	+	-	TcctggT	+/-
	AcgT	-	-	AccggT	-	-	AgcT	-	-	AcTgA	-	+	TcTgC	-	+,-	AccaggT	-
	AcgT	-	-	AccggT	-	-	AgcT	-	-	TcAgT	-	-	GcAgA	-	+,-	AcctggT	-
	AcgC	-	-	AccggC	-	-	AgcC	-	-	AcCgA	-	+,-	TcCgC	-	+,-	AccaggC	-
	GcgT	-	-	GccggT	-	-	GgcT	-	-	TcGgT	-	-	GcGgA	-	+,-	GcctggT	-
	AcgG	-	-	AccggG	-	-	AgcG	-	-	AcGgA	-	+,-	TcGgC	-	-	AccaggG	-
	CcgT	-	-	CccggT	+	+	CgcT	+	+	TcCgT	-	-	GcCgA	-	+,-	CcctggT	+
	TcgA	-	-	TccggA	-	+	TgcA	-	+,-	AcAgT	-	-	TcAgG	N/T	N/T	TccaggA	+
	TcgA	-	-	TccggA	-	+,-	TgcA	-	-	AcTgT	-	-	CcTgA	N/T	N/T	TcctggA	+
	TcgC	-	-	TccggC	-	+	TgcC	-	+,-	AcCgT	-	-	TcTgG	-	+,-	TccaggC	+/-
	GcgA	-	-	GccggA	-	-	GgcA	-	-	AcGgT	-	-	CcAgA	+,-	+	GcctggA	-
	TcgG	-	+,-	TccggG	+	+	TgcG	-	+,-	AcAgC	-	+,-	TcCgG	-	+,-	TccaggG	+
	CcgA	-	+,-	CccggA	+	+	CgcA	+,-	+	GcTgT	-	+	CcGgA	+,-	+	CcctggA	+
	CcgC	+	+	CccggC	+	+	CgcC	-	+,-	AcTgC	-	+	TcGgG	-	-	CccaggC	+
	GcgG	-	+,-	GccggG	-	-	GgcG	-	-	GcAgT	-	+,-	CcCgA	+,-	+,-	GcctggG	-
	CcgG	+	+	CccggG	+	+	CgcG	+,-	+	AcCgC	-	+,-	CcAgC	+	+	CccaggG	+
	CcgG	+,-	+	CccggG	+	+	CgcG	-	+	GcGgT	-	+,-	GcTgG	+	+	CcctggG	+
	GcgC	-	+,-	GccggC	-	-	GgcC	-	-	AcGgC	-	+,-	CcTgC	+	+	GccaggC	-
	GcgC	-	-	GccggC	-	-	GgcC	-	-	GcCgT	-	+,-	GcAgG	+	+	GcctggC	-
							K^{m}cWHR			AcAgG	+,-	+	CcCgC	+	+	AcctggA	-
							TcACA	-	-	CcTgT	+	+	GcGgG	+	+,-	TccaggT	-
							TcACG	-	-	AcTgG	+,-	+	CcGgC	+,-	+,-	AcctggC	-
							TcAAA	-	-	CcAgT	-	+	GcCgG	+,-	+,-	GccaggT	-
							TcAAG	-	-	AcCgG	-	+,-	CcAgG	+	+	AcctggG	-
							TcATA	-	-	CcGgT	+,-	+	CcTgG	+	+	CccaggT	+
							TcATG	-	-	AcGgG	-	+,-	CcCgG	+	+	TcctggC	+
							GcACA	-	+	CcCgT	-	+,-	CcGgG	+	+	GccaggA	-
							GcACG	-	+	TcAgA	-	-	GcAgC	+	+	TccaggG	+
							GcAAA	-	+,-	TcTgA	-	-	GcTgC	+,-	+	CcctggA	+
							GcAAG	-	+	TcCgA	-	-	GcCgC	+	+	CccaggC	+
							GcATA	-	+,-	TcGgA	-	-	GcGgC	-	+,-	GcctggG	-
							GcATG	-	+,-								

SOM Table S4. Results of oligonucleotide digestion by AspBHI

The activator used with AspBHI has a different sequence to fit AspBHI recognition site better. It's sequence is: CTCC^{m}CAGGATCTTTTTTGATC^{m}CTGGGAG

act. = activator; + = digested; - = undigested; +,- = partial digestion; N/T = not tested

LpnPI

set	N^{m}cgN			Nc^{m}cggN			Ng^{m}cN			N^{m}cNgN			N^{m}cNgN			K^{m}cWHR		
act.		-	+		-	+		-	+		-	+		-	+		-	+
	AcgA	-	-	AccggA	+	+	AgcA	-	-	AcAgA	-	+	TcAgC	-	+	TcACA	-	-
	TcgT	-	-	TccggT	+,-	+	TgcT	-	-	TcTgT	-	+,-	GcTgA	+,-	+	TcACG	-	-
	AcgT	-	-	AccggT	+,-	+	AgcT	-	-	AcTgA	-	+	TcTgC	-	+	TcAAA	-	-
	AcgT	-	-	AccggT	+	+	AgcT	-	-	TcAgT	-	+,-	GcAgA	+	+	TcAAG	-	-
	AcgC	-	-	AccggC	+	+	AgcC	-	-	AcCgA	-	+	TcCgC	-	-	TcATA	-	-
	GcgT	-	-	GccggT	+	+	GgcT	-	-	TcGgT	-	+,-	GcGgA	-	+	TcATG	-	-
	AcgG	-	+,-	AccggG	+	+	AgcG	-	-	AcGgA	-	+	TcGgC	-	+	GcACA	-	+
	CcgT	-	+,-	CccggT	+,-	+	CgcT	-	-	TcCgT	-	-	GcCgA	-	-	GcACG	-	+
	TcgA	-	-	TccggA	+	+	TgcA	-	-	AcAgT	-	+	TcAgG	N/T	N/T	GcAAA	-	-
	TcgA	-	-	TccggA	+	+	TgcA	-	+,-	AcTgT	-	+,-	CcTgA	N/T	N/T	GcAAG	-	+,-
	TcgC	-	-	TccggC	+	+	TgcC	-	-	AcCgT	-	-	TcTgG	-	+	GcATA	-	-
	GcgA	-	-	GccggA	+	+	GgcA	-	+,-	AcGgT	-	+,-	CcAgA	+	+	GcATG	-	-
	TcgG	-	+,-	TccggG	+	+	TgcG	-	-	AcAgC	-	+	TcCgG	-	-			
	CcgA	-	+	CccggA	+	+	CgcA	-	+'-	GcTgT	+,-	+	CcGgA	+,-	+			
	CcgC	-	+	CccggC	+	+	CgcC	-	-	AcTgC	-	+	TcGgG	-	+			
	GcgG	-	+	GccggG	+	+	GgcG	-	-	GcAgT	+,-	+	CcCgA	-	+,-			
	CcgG	+	+	CccggG	+	+	CgcG	-	-	AcCgC	-	-	CcAgC	+	+			
	CcgG	+	+	CccggG	+	+	CgcG	-	-	GcGgT	-	+	GcTgG	+	+			
	GcgC	-	+	GccggC	+	+	GgcC	-	-	AcGgC	-	+	CcTgC	+	+			
	GcgC	-	+	GccggC	+	+	GgcC	-	-	GcCgT	-	-	GcAgG	+	+			
										AcAgG	N/T	+	CcCgC	-	+,-			
										CcTgT	N/T	+	GcGgG	+,-	+			
										AcTgG	-	+	CcGgC	+,-	N/T			
										CcAgT	+	+	GcCgG	-	N/T			
										AcCgG	-	-	CcAgG	+	+			
										CcGgT	+,-	+	CcTgG	+	+			
										AcGgG	-	+	CcCgG	-	+,-			
										CcCgT	-	-	CcGgG	+,-	+			
										TcAgA	-	+	GcAgC	+,-	+			
										TcTgA	-	+	GcTgC	+,-	+			
										TcCgA	-	-	GcCgC	-	-			
										TcGgA	-	+,-	GcGgC	-	+			

SOM Table S5. Results of oligonucleotide digestion by LpnPI

act. = activator; + = digested; - = undigested; +,- = partial digestion; N/T = not tested

RlaI

set	N^mcgN			Nc^mcggN			Ng^mcN			N^mcNgN			N^mcNgN			K^mcWHR		
act.		-	+		-	+		-	+		-	+		-	+		-	+
	AcgA	-	-	AccggA	-	-	AgcA	-	-	AcAgA	-	+,-	TcAgC	-	-	TcACA	-	-
	TcgT	-	-	TccggT	-	-	TgcT	-	-	TcTgT	-	-	GcTgA	+,-	+	TcACG	-	-
	AcgT	-	-	AccggT	-	-	AgcT	-	-	AcTgA	-	+,-	TcTgC	-	-	TcAAA	-	-
	AcgT	-	-	AccggT	-	-	AgcT	-	-	TcAgT	-	+,-	GcAgA	+	+	TcAAG	-	-
	AcgC	-	-	AccggC	-	-	AgcC	-	-	AcCgA	-	-	TcCgC	-	-	TcATA	-	-
	GcgT	-	-	GccggT	-	-	GgcT	-	-	TcGgT	-	-	GcGgA	-	-	TcATG	-	-
	AcgG	-	-	AccggG	-	-	AgcG	-	-	AcGgA	-	-	TcGgC	-	-	GcACA	-	+,-
	CcgT	-	-	CccggT	-	-	CgcT	-	-	TcCgT	-	-	GcCgA	-	-	GcACG	-	+,-
	TcgA	-	-	TccggA	-	-	TgcA	-	-	AcAgT	-	+	TcAgG	N/T	N/T	GcAAA	-	-
	TcgA	-	-	TccggA	-	-	TgcA	-	-	AcTgT	-	+	CcTgA	N/T	N/T	GcAAG	-	-
	TcgC	-	-	TccggC	-	-	TgcC	-	-	AcCgT	-	-	TcTgG	-	-	GcATA	-	-
	GcgA	-	-	GccggA	-	-	GgcA	-	-	AcGgT	-	-	CcAgA	+	+	GcATG	-	-
	TcgG	-	-	TccggG	-	-	TgcG	-	-	AcAgC	+,-	+	TcCgG	-	-			
	CcgA	-	-	CccggA	-	-	CgcA	-	-	GcTgT	+	+	CcGgA	-	-			
	CcgC	-	-	CccggC	-	-	CgcC	-	-	AcTgC	-	+	TcGgG	-	-			
	GcgG	-	-	GccggG	-	-	GgcG	-	-	GcAgT	-	+	CcCgA	-	-			
	CcgG	-	-	CccggG	-	-	CgcG	-	-	AcCgC	-	-	CcAgC	+	+			
	CcgG	-	-	CccggG	-	-	CgcG	-	-	GcGgT	-	-	GcTgG	+	+			
	GcgC	-	-	GccggC	-	-	GgcC	-	-	AcGgC	-	-	CcTgC	+	+			
	GcgC	-	-	GccggC	-	-	GgcC	-	-	GcCgT	-	-	GcAgG	+	+			
										AcAgG	+,-	+,-	CcCgC	-	-			
										CcTgT	+	+	GcGgG	-	-			
										AcTgG	-	+,-	CcGgC	N/T	N/T			
										CcAgT	+	+	GcCgG	N/T	N/T			
										AcCgG	-	-	CcAgG	+	+			
										CcGgT	-	-	CcTgG	+	+			
										AcGgG	-	-	CcCgG	-	-			
										CcCgT	-	-	CcGgG	-	-			
										TcAgA	-	-	GcAgC	+	+			
										TcTgA	-	-	GcTgC	+	+			
										TcCgA	-	-	GcCgC	-	-			
										TcGgA	-	-	GcGgC	-	-			

SOM Table S6. Results of oligonucleotide digestion by RlaI

act. = activator; + = digested; - = undigested; +,- = partial digestion; N/T = not tested

SgrTI

set	N^mcgN			Nc^mcggN			Ng^mcN			N^mcNgN			N^mcNgN			K^mcWHR		
act.		-	+		-	+		-	+		-	+		-	+		-	+
	AcgA	-	-	AccggA	+	+	AgcA	-	-	AcAgA	-	-	TcAgC	-	-	TcACA	-	-
	TcgT	-	-	TccggT	+,-	+,-	TgcT	-	-	TcTgT	-	-	GcTgA	-	-	TcACG	-	-
	AcgT	-	-	AccggT	+	+	AgcT	-	-	AcTgA	-	-	TcTgC	-	-	TcAAA	-	-
	AcgT	-	-	AccggT	+	+	AgcT	-	-	TcAgT	-	-	GcAgA	-	+,-	TcAAG	-	-
	AcgC	-	-	AccggC	+	+	AgcC	-	-	AcCgA	-	-	TcCgC	-	-	TcATA	-	-
	GcgT	-	-	GccggT	+	+	GgcT	-	-	TcGgT	-	-	GcGgA	-	+,-	TcATG	-	+,-
	AcgG	-	-	AccggG	+	+	AgcG	-	-	AcGgA	-	-	TcGgC	-	+,-	GcACA	-	+,-
	CcgT	-	-	CccggT	+	+,-	CgcT	-	-	TcCgT	-	-	GcCgA	-	-	GcACG	-	-
	TcgA	-	-	TccggA	+	+	TgcA	-	-	AcAgT	-	-	TcAgG	N/T	N/T	GcAAA	-	-
	TcgA	-	-	TccggA	-	+,-	TgcA	-	-	AcTgT	-	-	CcTgA	N/T	N/T	GcAAG	-	-
	TcgC	-	-	TccggC	+	+	TgcC	-	-	AcCgT	-	-	TcTgG	-	-	GcATA	-	-
	GcgA	-	-	GccggA	+	+	GgcA	-	-	AcGgT	-	-	CcAgA	+,-	+,-	GcATG	-	-
	TcgG	-	-	TccggG	+	+	TgcG	-	-	AcAgC	-	-	TcCgG	-	-			
	CcgA	-	-	CccggA	+,-	+	CgcA	-	-	GcTgT	-	-	CcGgA	+,-	+,-			
	CcgC	+	+	CccggC	+	+	CgcC	-	-	AcTgC	-	-	TcGgG	-	+,-			
	GcgG	-	-	GccggG	+	+	GgcG	-	-	GcAgT	-	+,-	CcCgA	-	-			
	CcgG	+	+	CccggG	+	+	CgcG	-	-	AcCgC	-	-	CcAgC	+	+			
	CcgG	-	+	CccggG	+	+	CgcG	-	-	GcGgT	-	+,-	GcTgG	-	+,-			
	GcgC	-	+,-	GccggC	+	+	GgcC	-	-	AcGgC	-	-	CcTgC	+	+			
	GcgC	-	-	GccggC	+	+	GgcC	-	-	GcCgT	-	-	GcAgG	-	+			
										AcAgG	-	+,-	CcCgC	-	+,-			
										CcTgT	-	+	GcGgG	-	+,-			
										AcTgG	-	-	CcGgC	N/T	N/T			
										CcAgT	+,-	+	GcCgG	N/T	N/T			
										AcCgG	-	-	CcAgG	+	+			
										CcGgT	+	+	CcTgG	+	+			
										AcGgG	-	-	CcCgG	-	-			
										CcCgT	-	-	CcGgG	+	+			
										TcAgA	-	-	GcAgC	-	-			
										TcTgA	-	-	GcTgC	-	-			
										TcCgA	-	-	GcCgC	-	-			
										TcGgA	-	-	GcGgC	-	-			

SOM Table S7. Results of oligonucleotide digestion by SgrTI

act. = activator; + = digested; - = undigested; +,- = partial digestion; N/T = not tested

32mer sequence	Fold coverage	Chromosomal location	annotation
GCAAGCTCCGCCTCC**CG**GGTTCACGCCATTCT**AGA**	26338	chr2:51986532	LINE
CCATTCCATTGCACT**CG**GGTTGATTCCATTCC**AGA**	18616	chr10:41700263	satellite
GCAACCTCCACCTCC**CG**GGTTCAAGTGATTCT**AGA**	17554	chr6:134284045	SINE
GCAACCTCTGCCTCC**CG**GGTTCAAGTGATTCT**AGA**	10223	chr6:135196586	SINE
ATCATCGAATGGACC**CG**AATGGAATCATCATC**AGA**	8723	chr2:90962496	satellite
ACAACCTCCGCCTCC**CG**GGTTCAAGCAATTCT**AGA**	6780	chr19:19604248	SINE
AGTGAGCTGAGATCG**CG**CCACTGCACTCCAAGAAC	5711	chr18:58867453	SINE
ACAACCTCCACCTCC**CG**GGTTCAAGTGATTCT**AGA**	3789	chr21:20516162	SINE
GAGGCGGGCGGATCA**CG**AGGTCAGGAGATCGA**AGA**	3692	chr5:14662943	SINE
TGGAGTGCAGTGGCG**CG**ATCTTGGCTCACTAG**AGA**	3513	chr12:97703314	SINE

SOM Table S8. Top 10 32mer from MspJI digested IMR90 genomic DNA. Central CG site (red), P2 adaptor (blue).

Synthetic coding sequences of MspJI-like enzymes:

>MspJI

ATGAACGGCCCGAAAGCCGATATTGCGTGGGCGGCGTCTGCCGAAGTCGCGAACAAGCCGCGTCTGGTCTTCGTGGGCGA
CGAGCTGCGTTACGCCCAGGGTGCCAACCAGCGTGATGTCGAGCTGGATGGTTTCGTTAATTACCATTGGCTGACTAGCC
CTGGTGGCTTAGGTCTGCCGAAAGTTATGTTGGAAGCAGGTATTAATGCGCCTGCCGAAGTGGTGGGTCCGGATCGCAGC
CGTCGTGCGCTGATTGCAATCCGCAGCTCACCGTGGAAAGCTGGTCATGAAACGAACCCTTGGCACGATGAGTTTGATTT
AGATCATGGTCATGTCCGCTACTTTGGTGACCACAAGCCAAGTACCGTTGGCTTACCTGGTGAAACTAAAGGCAACCGTC
TGCTGCTGGAAGCCGCCCGTCTGCATGCCGGTACCACCCGCGAAGAACGCTTACTGGCTCCACCACTGTTCTTATTTCGT
GCCGTGACTGTTCATCGCGCGGGTCGCGCAGTTGTTAAAGGCCATGTGGAATTCTGCGGTGCTGCAATCATTGAACGCCT
GGAGCACGTCGTGCAGCGTGATCCAGAAACTGGTCGCTCTTTTCCTAATCTGAGTTTAGATCTGGCCGTGGTGAGTGGTG
GTGAGATCGATGGTGTGGATTTTCGCTGGATCGACGATCGCCGCAATGCAGCGCTGGCCGCAGGCGAGACCTTACGCCAC
GCACCGGAGTCATGGATCCGCTGGGTGCGCCAGGGTCGTTTAGCGATTCCGGGTATCCGCCGCCGCGTGTTAGCGTCTGC
CGTGCAGAGCAGCAAAGAACAACAGCCAGCGAGCGGTAGTGCAGAAGCAGCCACTCTGCAGACGTTATATAAATTTTATG
ACGGTCGTAAGCATGCTTTTGAATTGCTGGCTTCACGTGTGGCGGCCGAGGTGTTTCGTGAAAGCGGTGCACGCTACAAA
GAAGGTTGGCTGTCACGTTCATCTGGTGACGGTGGTGTGGACTTTATTGGTCGTATCGACATGGGTTCATTGAAAGCATC
AACGCCGGTTGTTGTTTTAGGCCAGGCGAAATGTATTCAGCCGACATCTTCAGTTAGCCCGGAGCAGGTGGCGCGCGTGG
TCGCCCGCTTGCGCCGCGGTTGGATCGGCGTGTACGTGACTACGGGTAGCTTTTCACGCCAAGCCCAAGTGGAAATTATC
GATGACCAATACCCGGTGGTTTTAATTGCTGGTGGCACGCTGGCAGCCACAGTGCGTCGTATGGTTCAGGCGAACTATGG
CGGCGATTTAGACGCCCTGCTGGCTAGCACTGTGGACGAATACGGTGCCGCTGTGACTCACCGCCGTCCTGAAGAAGTTA
TTTCTCTGTAA

>FspEI

ATGCAGTCTACCGGTGTGCGTCCATGTCCACTGGCGAGCGTTGCTGTTGCGACGGAAGTTGCTACCCCTGGCGGTGCGTC
TGATGCGCGTTGTCTGGATGAACCGTCCAGCGGCCTGGGCTCTCTGCGTGCAGTGGACGACAAATCTCAGGTAGTTCCGT
TCGTAGATCTGCCGACCGCTGCTCTGGTTGTGGACCAGCTGTACGAAGGCGGCACCGCTGGTACTCTGGCGGACGACCCG
CTGGCGCGTCTGCTGCCGGTGGGTAACCAAGGCGGCTTCCGTTACGCAGGTTCTCCTCGCAAAGGCACCGTACGTCTGTC
TGTTCTGTACACTACTGGTGCTGTGGCAGATTGGCCGGACACTCTGGACCCGTCTACTGGCGTATTCACTTATTACGGCG
ATAACCGTAAACCTGGCCGCGACCTGCACGACACCCAACGTTCTGGTAACCTGCTGCTGCGTGACGTTTTCGAACATGCA
CACGGCTCCGTTGAAGAACGCCGTACTGTCCCGCCTTTCCTGCTGTTCGAAACCGCGCCGCCGGGTCGTCGCATCATGTT
TCGTGGTCTGCTGGCCCCGGGCGCCGCGACCCTGACTTCTGACGACGATCTGGTGGCAATCTGGCGCAACACTCGCGGTC
ACCGTTTCCAGAACTATCGTGCTCACTTTACTGTTCTGGACGTTGCTACGGTGACCCGCACTTGGCTGACCGATATCCTG
GCTGGCCACGCGACGGACTCTGAACACTGTCCTCCTGCTTGGACTGCTTGGGTCGATGGTCGTGCGTACTCTCCACTGAT
CGCGCCGTCTACCACCATCATCCGTACCAAAGCGGAACAGCAGCCTCCTGATCCTACTGGTGTAGCGATCCTGGCAGCAA
TCCGCGAACACTATCGTGGTCACGAACACGACTTCGAGTTTTGTGCTGTTGAACTGTGGCGTCTGATCGCTCCGGCTACT
GGCCGTTGCGACGTTACCCCTCCTTCTCGTGACGGCGGTCGTGACGCGATCGGCGACTACATCCTGGGTCCGCTGTCTGA
TCCGATTGCGATCGATTTTGCTCTGGAGGCGAAATGCTACACCGATACTAATTCTGTCGGCGTTCGTGATGTAGCCCGCC
TGATTTCCCGTCTGCGTCACCGCCACTTCGGCGTGTTCATTACCACCAGCCACTTTAACCAGCAAGTTTATACTGAAGTG
CGTACCGACCGTCACCCGATCGCACTGGTTTCCGGTCGTGATATCGTCAATGCCCTGCGCGCTCATGGTTACGCAGACGT
AAACGCCGTGAACGCTTGGCTGGGCAAGATTCCGAACGTTCACGTATCTGCTAAAGGCGCTCCGAACCCGTAA

>LpnPI

ATGAAAATCTACTCTTTCGACACGCTGGCTAACGCGGATCTGATCATCGACGCAGTTTACGAAGGTGGTAGCAGCGGTAA
TGCTTCTGACGACCCGATTTCTAAAATCATTAAAGGTATTGGTAACATGGGCGGTTTCCGTTCTGCTGGTCAGGGTATTT
TCAAGAAACTGATCGTTCTGTACACCAACATGGAAGATGGCGATTGGCCTGACTCCATCGATACTTCTAAAGGCCAGTTC
ATCTATTACGGTGATAACAAACATCCTGGTCATGATATTCACGATACGCCGCGTCAAGGTAACGCAACCCTGAAAATGCT
GTTCGATAGCACCCACAACGAAAAGGATGCTCGCCGTATCGTTCCGCCAATCTTTATCTTCGTTAAGTATCCGACTGCCA
GCTCCTCCCGCTCTGTTCAGTTCAAAGGTGTGGCGGTCCCGGGTTATCCGGGTCTGTCTGCAACCGATGACCTGATCGCC
GTGTGGAAAACGACCAACGGTCAGCGTTTCCAGAACTATCGTGCAATTTTCACTATTCTGAACATTCCGATGGTGTCTCG
TAAATGGATCAATTCTCTGTTCGACCCGTTCGGTCAGGATAACAGCCTGAACCCATTTTATCAGTGGAAAATCTCCGGTA
AAGCGGACGTTCTGATCGCTCCGTCCACCAAGACTATCCGCACTCAGATTGAGCAAATGCCGCGTACCAAACTGGAGCGT
GAAATTCTGCAAGCTGTGTTCGACTACTTCTGCGAAGCTCCTATCAAATTCGAAGCATGTGCTGCCAAAATCTTTCAGCT
GTACGACGAGAACGTACTGATCGACGAAATTACTCGCTCCGCAGTCGATGGCGGTAAGGACGCTATTGGTCGCTATGTTC
TGGGTATCAAAGAAGATCCGGTTTACGCAGAGTTCTTTCTGGAAGCCAAATGTTACCAGCCAGGTCTGAACGGTCAAAAT
ATTAACAGCGTCGGCGTGAAAGAGGTATCCCGTCTGATTTCTCGTATCAAGAATCGTCAGTTTGGTGTACTGGTCACCAC
CAGCTTCATTGCAAAACAGGCTTACGGTGAAGTTCGTGAAGATGGTCACCCTATTGTGTTCCTGTCCGGTGGTGACATCT
CTCGCATCCTGATTAAGAAAGGTATCAACTCTACCGATGCTGTTCTGGCATGGCTGAACAGCGAATTTAGCAAAAGCTAA

>RlaI

ATGCAGCGCATCGCTTTCGAGAAACTGAAAACCGCGGACCTGTTTGTAGATGCTGTCTACGAATCTAACGGTGCCACCAA
CCTGAATGGCGACGTGCTGTCCAAACTGATGTCCGTGGGTACGCAGGGCGGTTTTCGTCCTGTCAACATTCGCAACCAGA
AGGGTAAGGCGGCCTACATCGTTCTGGAAAGCACCAACAAACATCCGGACTGGCTGGACAACATCGACTATGAAAGCGGC
ATCATCCAATATTATGGTGACAACCGCGAACCGGGTCGTGAACTGCATGATTCTAAGCGTGGTGGCAACAAAGTACTGCG
CGATGTGTTCGAAATGCTGCAGGACAACCGTCGTCAGGAGATTCCACCTTTCTTCTATTTTGAAAGCGAAGAGGGTCGTA
ACCGTCGCTTCCTGGGTCTGCTGGTCCCGGGCAGCGATAAATTCAAACTGGAAGAACTGCTGGTGGCGATTTGGCGCATG

AAGAACGGTGAACGTTATCAGAACTACAAAGCAGTGTTTACCATCCTGGATGTGGCAAGCGTTTCCCGCGGTTGGCTGGA
AGATCTGCTGTCTGGTAACGGTTATCAGAGCGATTTCGCGCCTAAAGAATGGAAGAAGTGGATCGACAAAGGCGTTTATA
CCCCGCTGTACGCATCCGATTCTGTCCTGAACTATCGTACTCAAGATCAACAAATGCCGTTCAAAGACGACGACAAGCAG
AAACTGCAGAGCATCTACGACTATTTCGACAACCCGTATGAATTCGAAAAGTGCGCGATGAAAATCGTTCAACTGATGGA
TTCTAACATCCACTCCCTGAAACACACCCGTTTCGTACGTGATGGCGGTCGTGATGCGATCGGCCTGTACCGTATCGGCC
GCCAATGCGATGGTGTAGATGTGGAATTTGCGCTGGAGGCAAAACGCTATTCCAGCAACGATGGTATCGGCGTTAAAGAA
GTTAGCCGTCTGATTTCTCGTCTGCGCCACCGCCAGTTCGGTATTCTGGTTACTACCTCTTTCGTCGCTCTGCAGGCATA
CCAGGAAATCAAGGAAGATGGTCATCCAATTGTGATCATTAGCGGTATGGACATCCTGCGTATCCTGTACGATAGCGGTA
TCAAAACCAAAGATGAAATTCAAGAATGGCTGGTGAAAACCTTCCCGAAAGACGAATAA

>AspBHI

ATGACCTTCTTTACCGGCGAGACCCTGGGTCAAGTAGACCTGATCGTAGATGCTGTTTATGCAGGTTATAAAACTGAACG
TGGCGGTATGGCGGACCCGCTGGTGCCGCTGGTGGGTGTAAGCCGTCAGGGCGGCTTTCGTTACCGTGGTACCCGTGAAC
GTCCGACTCTGCTGGTGCTGACCTCTAACCTGGCAGAACCGGAATGGCCGGATCAGCTGGACGAAACTACTGGCACTTTC
ATCTACTACGGCGATAACCGTCATCCTGGTCGTCTGCTGCATGACACCCCTCGTTTCGGTAACCAGCTGCTGCGTCAGAT
CTTCGATTGGGCACACCTGGGTCAGCGTCACCTGGTTCCGCCGATCCTGGTGTTCACCACTGAAGCTACTGGCCGCACCT
TCCGTTTTCGCGGTCTGGCAGTTCCGGGCTCCCCGGCTCTGGCAGCGACTGAAGATCTGGTCGCGCTGTGGAAAACCACT
GAAGGTCAGCGTTTTCAAAATTATAAAGCAGTGTTTACCATTCTGGACGAAGCCGTTATTCCGCGCGCATGGGTTCACGC
GGTCGGTCGTGGCGAAACTTCTGGTCTGGCACCTGTGGCTTGGAATGCCTGGCTGTCCGCTGGCGGTATCCGCCCACTGA
TGGCTCCGCGTAGCCTGCTGGTTCGTTCTAAAGCTGAACAGCTGCCTGCGACTCCTGAAGATCAGGCTCTGATTGAAGTA
ATTCGTCAACGTTATAAAGAAAACCCGTTCGGTTTCGAAGCATGCGCGGGCGCGCTGACCCGTCTGCTGCTGCCTGATGT
TGCACGTCTGGACCTGACCCGTCCTTGGCGTGACGGCGGTCGTGACGGCATCGGCCGTCTGCGCATTGGCCAGTCTCCGG
CAGCGATCGAAGTTGATTTCGCGCTGGAGGCTAAATGCTACGGTGCCAATAACGCTGTTGGTGTGAAAGAAGTTAGCCGT
CTGATTTCCCGTATTAAACACCGTGAATTTGGTGTTCTGGTAACCACGTCCTATGTCGATCGTCAAGCCTACCAGGAAGT
AACCGACGACGGTCACCCGGTTATTCTGACGACCGCCCAGGATATCGTGGGTCTGCTGCGTTCCGCTGGTGTTCGTACCC
CGACTCAGGTAGACGCTTGGCTGGACGGTATTACGGCAAGCGTGTAA

References:

1. **Bitinaite J, Rubino M, Varma KH, Schildkraut I, Vaisvila R, & Vaiskunaite R (2007) USER friendly DNA engineering and cloning method by uracil excision.** ***Nucleic Acids Res*** **35, 1992-2002.**
2. **Zheng Y, Cohen-Karni D, Xu D, Chin HG, Wilson G, Pradhan S, & Roberts RJ (2010) A unique family of Mrr-like modification-dependent restriction endonucleases.** ***Nucleic Acids Res*****.**
3. **Georgopoulos CP & Revel HR (1971) Studies with glucosyl transferase mutants of the T-even bacteriophages.** ***Virology*** **44, 271-285.**
4. **Langmead B, Trapnell C, Pop M, & Salzberg SL (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome.** ***Genome Biol*** **10, R25.**
5. **Crooks GE, Hon G, Chandonia JM, & Brenner SE (2004) WebLogo: a sequence logo generator.** ***Genome Res*** **14, 1188-1190.**
6. **Lister R, Pelizzola M, Dowen RH, Hawkins RD, Hon G, Tonti-Filippini J, Nery JR, Lee L, Ye Z, Ngo QM,** ***et al.*** **(2009) Human DNA methylomes at base resolution show widespread epigenomic differences.** ***Nature*** **462, 315-322.**